Statistics for Electronic Resources

by
Caroline Elizabeth Dean
DNXCAR003

A minor dissertation submitted in partial fulfillment of the requirements
for the award of the degree of
Master of Library and Information Science

Faculty of the Humanities
University of Cape Town
2008

DECLARATION

This work has not been previously submitted in whole, or in part, for the award of any degree. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works, of other people has been attributed, and has been cited and referenced.

Signature: Date: 27 May 2008

Acknowledgements

My thanks to Associate Professor Karin de Jager for her unending patience and valuable input.

Thank you too, to family, friends and colleagues for the encouragement to persevere and the offers of assistance. It is finally finished.

All Praise be to the Lord Jesus Christ.

Abstract

Electronic resources represent a large portion of many libraries' information resources in the current climate of hybrid libraries where print and electronic formats coexist. Since the dramatic uptake of electronic resources in libraries during the 1990's the topic of usage statistics has been on librarians' lips. The expectations that librarians had of being able to compare resources based on usage statistics were soon dashed as it became apparent that electronic resource providers were not measuring usage uniformly. Given the initial disappointments that librarians had in terms of electronic resource usage statistics the author set out to find the reasons why librarians were keeping statistics for electronic resources, which statistics they were keeping for electronic resources, and what were the issues and concerns with regard to statistics for electronic resources.

To get an international answer to these questions a literature review was undertaken. The South African point of view was sought through an e-mail survey that was sent out to the 23 South African academic libraries that form the South African National Library and Information Consortium (SANLiC). A 65% response rate was recorded.

The international and South African answers to the three questions were very similar. The study found that the reasons why librarians keep electronic resources statistics were to "assess the value of different online products/services"; to "make better-informed purchasing decisions"; to "plan infrastructure and allocation of resources"; and to "support internal marketing and promotion of library services". The study also found that the statistics that librarians were keeping are: sessions, searches, documents downloaded, turnaways, location of use, number of electronic resources, expenditure and virtual visits. The number of virtual visits was kept by international libraries but no South African libraries reported keeping this information. The concerns that were raised by both international and South African libraries were found to be about: the continued lack of standardisation; the time-consuming nature of data collection; the reliability of the usage data; the fact that the data need to be looked at in context; the management of the data;

and how to count electronic resources. Clear definitions of the latter are essential. A concern raised in South Africa but not in the international literature is that there exists a lack of understanding amongst some South African librarians of the basic concepts of electronic resources usage statistics.

The author concludes with a suggestion that the CHELSA Measures for Quality be implemented so that librarians can see that the collection of usage data for electronic resources has some purpose. Once this is in place one or more training events under the auspices of SANLiC should be organised in order to train librarians in the best practice of electronic resource usage statistics.

Abbreviations and Acronyms

AIP	American Institute of Physics
ARL	Association of Research Libraries
CAUL	Council of Australian University Librarians
CD-ROM	Compact Disc, Read Only Memory
CHELSA	Committee for Higher Education Librarians of South Africa
COSALC	Coalition of South African Library Consortia
COUNTER	Counting Online Usage of Networked Electronic Resources
CSIR	Council for Scientific and Industrial Research
DVD	Digital Versatile Disc or Digital Video Disc
ERMS	Electronic Resource Management System
HEI libraries	Higher Education Institution libraries
HTML	Hypertext Markup Language
ICOLC	International Coalition of Library Consortia
ILS	Integrated Library System
IOP	Institute of Physics
IP	Internet Protocol
ISI	Institute for Scientific Information
ISO	International Standards Organisation
ISSN	International Standard Serial Number
IT	Information Technology
JCR	Journal Citation Reports
MPS	Macmillan Information Processing Services
NISO	National Information Standards Organization
URL	Uniform Resource Locator
PALS	Publisher and Librarian Solutions
PDF	Portable Document Format
RSC	Royal Society of Chemistry
RSS	Really Simple Syndication

SANLiC	South African National Library and Information Consortium
SASLI	South African Site Licensing Initiative
SCONUL	Society of College, National and University Libraries
SFX	Special Effects
UCT	University of Cape Town
UKSG	United Kingdom Serials Group
USA	United States of America
XML	Extensible Markup Language

Table of Contents

List of tables

List of Figures

Chapter One: Introduction

1.1 Introduction

During the long tradition of information being available from libraries in a physical format and libraries being physical places to visit, statistics have been used by libraries to measure their services. Over the years measurement standards were established so that libraries knew what to assess and how to do it (White & Kamal, 2006: 22). The usage of print collections has mainly been measured by keeping track of circulation figures. In the print environment records have been kept of expenditure on resources and the number of volumes on the shelves. Libraries keep a count of the number of people entering their doors and the number of reference queries librarians have to answer.

Libraries no longer operate in a purely physical environment, but more often in a hybrid environment with print and electronic information resources side-by-side. With the development of the internet and the wealth of electronic resources available, large portions of libraries' collections are no longer available on shelves in the library. They are available electronically on CD-ROM (compact disc, read-only memory), DVD (digital video disc) or the internet. In order to fully report on their activities libraries need to keep statistics for their electronic resources.

1.2 Statistics for electronic resources

The 1990's saw the start of a dramatic uptake of electronic resources in libraries (Blecic, Fiscella & Wiberley, 2007: 26). Librarians are now at the stage where they have to decide what metrics to use for the measurement and evaluation of electronic resources and how to use them (Bertot, et al, 2004: 30).

Since the introduction of electronic resources in libraries, it has been generally recognised that it should be possible to obtain usage statistics from the computer systems that facilitate access to the information in the resource (Cox, 2003/2004: 92; Sack, 2003/2004: 36). Librarians envisaged that this usage data would enable them to see how library material was being used, more easily than this was ever done in the print environment (Gallagher, Bauer & Dollar, 2005: 172). Unfortunately, in the very early days of usage statistics for electronic resources librarians were disappointed, primarily as the ways in which publishers obtained and reported the usage data were not standardized. This meant that librarians were not clear about what the usage data actually meant and comparisons between resources were not possible as usage was being measured differently from resource to resource.

Although the situation with regard to usage statistics has not yet been fully resolved, great strides have been made in terms of standardising the measurement of usage of electronic resources.

1.3 Objective of research project

The aim of this research project is to establish what libraries across the world are doing in terms of statistics for electronic resources and to establish where South African libraries stand in relation to the international situation.

1.4 Research questions

The following research questions are addressed in this research project:

1. Why are libraries keeping statistics for electronic resources?
2. Which statistics are libraries keeping for electronic resources?
3. What are the issues and concerns with regards to statistics for electronic resources?

1.5 Research methodology

In order to learn what libraries internationally are doing with regards to statistics for electronic resources a literature review was conducted to seek answers to the research questions listed on the previous page.

Based on the findings from the literature review a questionnaire (see Appendix, p.75) was constructed and distributed by e-mail to the libraries that form SANLiC (South African National Library and Information Consortium; see discussion on p.36). The aim of the survey was to learn what the libraries of the South African tertiary academic institutions are doing with statistics for electronic resources in comparison to their international counterparts.

The South African libraries that are members of SANLiC were chosen as the target population as this is a clearly defined group of libraries in which electronic resources form a vital source of information.

1.6 Clarification of terms

Most terms used in this research project that might require clarification will be explained when they are used. The terms covered here are those used in the title of the dissertation and throughout the work.

Electronic resource has been used to encompass all information sources (books, journals, databases, and other formats) that exist in an electronic/digital format, be that computer disc, CD-ROM, DVD or accessed on the internet. When the author uses the terms *Electronic journal* or *database,* reference is being made to that particular type of information source.

The term *database* is used to refer to indexing and abstracting databases or the so-called aggregated databases which are essentially collections of full-text electronic journals.

An *electronic journal* is a journal that exists in digital form on the internet, or on CD-ROM or DVD. It can be either an electronic version of a print journal, or it can exist solely in electronic form.

For the most part the statistics focused on in this research project are usage statistics as these are the statistics that are most prevalent. Other statistical data would be information such as expenditure on resources and numbers of titles.

1.7 Outline of chapters

Chapter One has provided an introduction to the research project.

Chapter Two will discuss the types of statistics available for electronic resources and the various sources of this information.

Chapter Three is a brief discussion of some of the projects, initiatives and tools that have bearing on statistics for electronic resources.

Chapter Four seeks answers to the research questions posed on page 2 from the international perspective through a literature review.

Chapter Five briefly describes the South African academic library environment as an introduction to the following chapter.

Chapter Six covers the survey of South African academic libraries that was conducted in order to establish where these libraries stand in terms of statistics for electronic resources.

Chapter Seven provides a summary of the preceding chapters with some suggestions for the way forward.

Chapter Two: Types and sources of statistics for electronic resources

2.1 Introduction

The first section of this chapter will provide a summary of the types of statistical data available for electronic resources. The second section lists the various sources where librarians can obtain this information. Both sections aim to provide background information for the research questions posed in Chapter One.

2.2 Types of statistical data for electronic resources

The following section will discuss the types of statistics that are available for electronic resources. These are usage statistics, numbers of resources and expenditure.

2.2.1 Usage statistics

Usage statistics form the focus of this research project – why libraries keep them, which data they are keeping, and what their concerns are regarding them. The term *usage statistics* includes information such as:

- number of full-text articles downloaded from an electronic journal or database/collection of electronic journals;
- number of searches performed on a database, or platform (a platform is website that hosts several databases);
- number of sessions, or successful connections, to a database or platform;
- number of turnaways, or rejected sessions, for an electronic journal or database.

2.2.2 Number of resources

Number of resources refers to the number of electronic resources held by a library. This information could be broken down by format, for example number of databases, number of electronic journals, and number of electronic books. These numbers could be further broken down by subject or subscribing department.

In addition to the subscribed resources, a library might keep track of the number of free electronic resources to which they provide links from their library home page or list of electronic journals.

2.2.3 Expenditure

The last type of statistical data for electronic resources is expenditure. This includes information such as total amount spent on electronic resources or the percentage of budget committed to electronic resources. This information could also be broken down by format, subject, or subscribing department.

2.3 Sources of statistical data for electronic resources

Librarians can access statistical data for electronic resources from a variety of sources, as explained below.

2.3.1 Publisher systems

Usage data is often made available to librarians through the administration functionality of each resource on the publisher's website. The librarian is provided with a username and password to log into the administration features and access the usage data.

Alternatively, some publishers send files of usage data to librarians via e-mail. At present, all publishers do not yet supply their usage data to libraries in the standardised format according to the international code of practice known as Project COUNTER. (See p.13 for a full discussion of Project COUNTER.)

Although a time-consuming exercise, usage statistics for electronic resources are relatively easy to gather from websites. They can demonstrate a level of actual use that was never determinable from the circulation statistics gathered from libraries' integrated systems, as one can often see usage of specific journal titles within big databases (Blake & Schleper, 2004: 461).

2.3.2 Electronic resource management tools

Electronic resource management tools such as Open URL resolvers (for example SFX) and alphabetical electronic journal list providers (for example, SerialsSolutions, EBSCO AtoZ, TDNet Journal Manager) can provide aspects of usage statistics for electronic resources.

SerialsSolutions, an electronic journal list provider, can provide data on the number of times each electronic journal is selected from the list, a so-called "click-through". Although this information can provide an indication of the level of use being made of each electronic journal, it does not give the complete picture as it is not counting the use made of the journal by people accessing the journal website directly.

SFX, an Open-URL resolver, can also provide the "click-through" usage data described in the previous paragraph. (An Open-URL resolver is essentially a system that facilitates linking between resources (Emery, 2005:139), typically from an indexing database to a full-text electronic journal.)

These tools are typically also able to supply data on the number of electronic journals available to library users. This is particularly useful when it comes to establishing the number of titles accessible through aggregated databases. (These are databases of mainly full-text journal content from a variety of publishers, assembled into a database by a third party publisher. Frequently, there is an embargo on the latest issues of journals in these databases.)

2.3.3 Library management systems/Integrated library systems (ILS)

Typically the acquisitions/serials module of the ILS (or equivalent system) can provide data on the number of subscriptions held and the expenditure made on these subscriptions.

2.3.4 Web server logs

Institutional Information Technology (IT) departments keep an automatic log, or record, of websites accessed by internet users in the institution. Libraries can estimate the number of sessions for a resource by looking at institutional web server logs, however this data cannot supply information on searches and article downloads (Blecic, Fiscella & Wiberley, 2001: 435). Another shortcoming of using web server log data to count database usage is that if users bypass the front screen of the database, and link through to an individual journal within the database, the session would not be counted towards that database's use as the web server log would show access being made to a different website or URL (Coombs, 2005: 601). This dissertation will not look at usage data from web server logs, choosing rather to focus on the usage data provided by publisher systems as this is the data stipulated by COUNTER. (see p.15)

2.3.5 Manual count

Where there is no automated alternative, librarians have resorted to manually counting use. This is an unreliable method of counting as it relies on librarians being able to watch what is happening at the resource and manually counting as each user sits down to use it. A manual session count would be used to tally the use made of stand-alone databases, particularly CD-ROM/DVD resources, actually in the libraries.

2.4 Summary of Chapter Two

Chapter Two was a brief overview of the types of statistical data available for electronic resources and the sources of this information. The types of data covered were usage statistics, numbers of resources and expenditure. Depending on their type, these data elements are available from publisher websites, electronic resource management tools, library management systems, web logs or through a manual count.

The following chapter discusses some of the projects, initiatives and tools that are relevant in a discussion about statistics for electronic resources and which are required for an understanding of the importance of these statistics in libraries.

Chapter Three: Projects, initiatives and tools

3.1 Introduction

This chapter will not attempt to cover in detail all the projects and initiatives focused on electronic resource measurement that have taken place. Rather, it will focus on two of the more influential of these: the ICOLC *Guidelines for statistical measures of usage of web-based information resources*, and Project COUNTER, in order to provide background information on the electronic resources statistics environment for the research questions posed in Chapter One.

The third and fourth sections of this chapter discuss what the author has chosen to call "tools". The purpose of ScholarlyStats and SUSHI are to help librarians with the management of statistics for electronic resources.

The final section of the chapter looks at a new project that is seeking to find a new measure of journal quality based on usage statistics.

3.2 Projects, initiatives and tools

A few electronic resource statistics projects, initiatives and tools will be discussed in the following section.

3.2.1 ICOLC Guidelines

The term *e-metrics* refers to the measurement of "the activity and use of networked information" (White & Kamal, 2006: 5). The ICOLC Guidelines have proved to be one of the most influential outcomes from the various e-metrics projects that have been undertaken in the library environment (White & Kamal, 2006:24).

The International Coalition of Library Consortia (ICOLC) was formed in 1996. The Coalition is a group of 200 library consortia from all over the world who get together to address issues of mutual interest. In November 1998 the Group issued the first version of their *Guidelines for statistical measures of usage of web-based information resources.* The purpose of the Guidelines was to provide the Group members with the information they required in order to ensure that all their consortium members were receiving usage data for their licensed resources. The Guidelines were also meant to provide publishers with information that would enable them to understand their customers' needs. The Guidelines were revised in December 2001 and then again in September 2006. (ICOLC, 2006) They are available at http://www.library.yale.edu/consortia/webstats06.htm [2008, January 28]

ICOLC's motivation in compiling its Guidelines was to meet the interests of its members "by defining and creating a common set of basic use information requirements that are an integral and necessary part of any electronic product offering". The Guidelines define the following aspects of usage reporting:

1. Minimum requirements
2. Privacy and user confidentiality
3. Institutional or consortial confidentiality
4. Access
5. Delivery
6. Definitions
7. Report formats

(ICOLC, 2006)

The *minimum requirements* of the ICOLC Guidelines are that the following "data elements" have to be made available according to database, institutional IP addresses, by consortium and by time period:

- "number of sessions (logins)",
- "number of queries (searches)",
- "number of menu selections",

- "number of full-content units examined, downloaded, or otherwise supplied ...", and
- "number of turnaways".

The requirement of *privacy and user confidentiality* stipulates that usage data that reveal any personal information will remain confidential. *Institutional or consortial confidentiality* is also required. Providers may not make available to others any statistics about institutions or consortia without permission. What is permitted, is that these statistical data may be made available as "part of an aggregate grouping of similar institutions for the purposes of comparison." The aggregated database JSTOR makes this sort of information available (Nisonger, 2000: 301).

The Guidelines stipulate that consortium administrators should have *access* to the usage data for their consortium members and consortium members should have access to the data of all other consortium members. It is stipulated that usage reports should be *delivered* on a "web-based reporting system preferably on a real time basis, but at least within 15 days after the end of the month" (ICOLC, 2006).

The ICOLC Guidelines require providers to *define* each of the data elements they supply in specific *report formats*. The Guidelines also provide sample reports as illustrations for providers of the minimum requirements (ICOLC, 2006).

The introduction to the new (2006) revision of the ICOLC Guidelines looks to future and additional standards: "With the continuing endorsement of 83 consortia from around the world ..., this revision reflects ICOLC's previous endorsement of Project COUNTER and the ICOLC community's new endorsement of NISO's Standardized Usage Statistics Harvesting Initiative (SUSHI) protocol and reliance on XML as the standard delivery format for usage statistics" (ICOLC, 2006).

3.2.2 COUNTER

As a result of librarians' stated need for uniformity in terms of usage statistics, Project COUNTER (Counting Online Usage of NeTworked Electronic Resources) was developed. It began, in 2002 (White & Kamal, 2006:22), as an initiative of the PALS (Publisher and Librarian Solutions) group in the United Kingdom and is now the international standard (Pesch, 2004a: 4; Shepherd, 2006: 144) for usage statistics for electronic resources. COUNTER is owned by its members who are "publishers, intermediaries, libraries, consortia and industry organisations" (Shepherd, 2006: 151).

The Project's objective was to "develop agreed international Codes of Practices governing the recording and exchange of online usage data for different categories of content." Codes of Practice have been released for journals and databases (release 1 in January 2003; release 2 in April 2005) and books and reference works (in 2006) (Shepherd, 2006: 143). COUNTER specifies how and when usage should be counted thereby providing consistency across publishers (Pesch, 2006: 150).

The full text versions of the Codes of Practice are freely available on the COUNTER web site (www.projectcounter.org). Whereas the standards created by NISO and ISO are formal, the COUNTER Code of Practice is a "voluntary set of guidelines that address terminology, layout and format of the report, processing of usage data, what categories or filters should be available, and delivery of reports" (Pesch, 2006: 148).

The usage reports that are specified in Release Two of the *COUNTER code of practice for journals and databases* for COUNTER compliancy are listed in Table 1. In order to become "COUNTER-compliant" publishers must supply their customers with the usage reports that are applicable to their product (COUNTER, 2005a: 9). The *Code of Practice* clearly defines all the data elements that are required in each report and provides an example of each report demonstrating the layout of the report.

Table 1: COUNTER usage reports for journals and databases

Journal Report 1 (JR1)	Number of successful full-text article requests by month and journal
Journal Report 2 (JR2)	Turnaways by month and journal
Database Report 1 (DB1)	Total searches and sessions by month and database
Database Report 2 (DB2)	Turnaways by month and database
Database Report 3 (DB3)	Total searches and sessions by month and service

JR1 and JR2 both list full journal names, print ISSNs (International Standard Serials Number) and electronic ISSNs against the number of successful full-text article requests and number of turnaways respectively, per month (COUNTER, 2005a: 7, 12). DB1 and DB2 show searches and sessions, and turnaways, respectively, by month and database (COUNTER, 2005a: 14, 17). DB3 reports on searches and sessions by month and service (COUNTER, 2005a: 19). The COUNTER definition of a *service* is

> *A branded group of online information products from one or more vendors that can be subscribed to/licensed and searched as a complete collection, or at a lower level*
>
> (COUNTER, 200b: 9).

One example of this is ScienceDirect where one can search each journal on the platform individually, or one can search the platform as a database.

Optional Additional Usage Reports are listed in Appendix H (available at http://projectcounter.org/r2/R2_Appendix_H_Jan_08.doc) [2008, January 30]
of the *Code of practice for journals and databases.* These are not required for COUNTER-compliancy. Journal Report 3 reports the "Number of Successful Item Requests and Turnaways by Month, Journal and Page-Type" and Journal Report 4, the "Total Searches Run by Month and Service" (COUNTER, 2005c).

As with the *ICOLC Guidelines*, the *COUNTER code of practice for journals and databases* has specific requirements in terms of report delivery:

> - "Reports must be provided either as a Microsoft Excel file, CSV file, or as a file that can be easily imported into Microsoft Excel pivot tables. In addition reports may also be provided in XML format.
> - Each report should reside in a separate file or page to avoid files of unwieldy size
> - Reports should be made available on a password-controlled website (accompanied an e-mail alert when data is updated). Access to consortia level reports must be through the same user id and password for all consortium members. (This user id password must be different from those used for administrative purposes for each institution.)
> - Reports must be readily available
> - Reports must be provided monthly
> - Data must be updated within four weeks of the end of the reporting period
> - All of last calendar year's data and this calendar year's to date must be supplied"
>
> (COUNTER, 2005a: 22)

When a database or journal is purchased by a consortium, the COUNTER reports that need to be supplied to customers are JR1 and DB1. These reports need to show usage for the consortium as a whole, but also need to break down the usage per institution (COUNTER, 2005a: 21-22).

The COUNTER code of practice also sets up protocols that define at what point usage via aggregated databases and gateways should be counted. For example, for "referral from an aggregator or gateway" it is the responsibility of the vendor or publisher to supply the COUNTER usage statistics. This applies to services such as electronic journal list providers that send users to the web sites of publishers in order for them to access the full-text articles for the electronic journals. Although the electronic journal list provider can provide usage for the service, the "COUNTER usage" for the electronic journals is counted at the publishers' web sites (COUNTER, 2005a:24-25).

Appendix D (available at http://projectcounter.org/r2/R2_Appendix_D_Guidelines_for_Implementation.doc) [2008, January 30] of Release Two of the COUNTER *code of practice for journals and databases* provides guidelines for publishers on how the Code should be implemented. Depending on their usage counting practices before implementing the Code of Practice,

publishers might see a decrease in usage statistics after implementation. Reasons for this include:

(1) "The implementation of time filters to eliminate multiple counting of requests when these result in only one download of a document ...";

(2) "The specification of the http return codes that are deemed to identify a 'successful request";

(3) "The standardisation of definitions of 'successful request' and 'full-text' and many other terms" (Gillingham, 2004).

The *COUNTER Code of Practice for Books and Reference Works: Release 1* (available at http://wwwprojectcounter.org/cop/books/cop_books_ref.pdf) [2008, January 29] was released in March 2006. This *Code of Practice* does not define a *book*, but provides the following definition of a *reference work:*

> *An authoritative source of information about a subject: used to find quick answers to questions.*
>
> *Examples: Dictionary, encyclopedia, directory, manual, guide, atlas, bibliography, index*
>
> (COUNTER, 2006: 10)

The structure and requirements for COUNTER-compliancy for books and reference works are essentially the same as those required for journals and databases; however the required usage reports are different. These reports are listed in Table 2, on the following page. As for journals and databases, in order to be COUNTER-compliant for books and reference works publishers must provide the reports that are appropriate for their product (COUNTER, 2006: 12).

Table 2: COUNTER usage reports for books and reference works

Book Report 1	Number of Successful Title Requests by Month and Title
Book Report 2	Number of Successful Section Requests by Month and Title
Book Report 3	Turnaways by Month and Title
Book Report 4	Turnaways by Month and Service
Book Report 5	Total Searches and Sessions by Month and Title
Book Report 6	Total Searches and Sessions by Month and Service

(COUNTER, 2006: 13-30)

The definition of a *section* for Book Report 2 is

A subdivision of a book or reference work.

Examples: chapter; entry

(COUNTER, 2006:10)

The definition of an *entry is*

A record of information in some categories of reference work.

Example: a dictionary definition

(COUNTER, 2006: 9)

The definition of a *service* is available on page 14.

As more publishers participate in the COUNTER project there is now greater standardisation in the way that usage is reported (Blake & Schleper, 2004: 461; Soule, 2006). In order to be COUNTER-compliant publishers are required to have their usage reports audited annually by a suitably qualified auditor (COUNTER, 2005a: 23). This makes for more reliable usage statistics and for usage that can be compared across resources (Gillingham, 2004). This is exactly what librarians are seeking and they are increasingly insisting that vendors supply COUNTER-compliant statistics.

3.2.3 ScholarlyStats

ScholarlyStats was created by MPS Technologies. This service collates COUNTER reports from various publisher and vendor sites and makes them accessible to the

subscribing library in one location, also making some analytical functionality available (Caldwell, 2006a: 4).

There are different subscription levels for the service based on the number of resources/platforms that ScholarlyStats will manage for the subscriber. The subscriber provides ScholarlyStats with the usernames and passwords required for each platform in order to sign into the administrative area where the usage statistics are accessible. Instead of the library having to sign into each platform on a monthly basis in order to access the usage data, ScholarlyStats does this on behalf of the library. Each month an e-mail is sent to the subscribed libraries informing them that their statistics have been collated for the month and are accessible on the ScholarlyStats web site, http://www.scholarlystats.com [2008, January 29]. Each subscribed library is provided with a unique username and password which allows them to access the collated data for their institution.

3.2.4 SUSHI

The Standardized Usage Statistics Harvesting Initiative (SUSHI) was conceived to develop an "automated protocol for moving COUNTER XML reports from providers to libraries" (Chandler & Jewel, 2006a: 82). Full details and the draft standard are available online at http://www.niso.org/committees/SUSHI/SUSHI_comm.html [2008, January 28].

The initial development group consisted of representatives from Ex Libris, Innovative Interfaces, Swets Subscription Services, EBSCO Subscription Services, the California Digital Library and two librarians from the United States of America (Chandler & Jewel, 2006b: 69). This protocol should make the task of tracking usage easier for librarians as processes become more automated (Caldwell, 2006b: 3). Once functioning, SUSHI could be used by ScholarlyStats and Electronic Resource Management Systems to harvest usage statistics from compliant publishers (Caldwell, 2006a:4; Chandler & Jewel, 2006b: 69).

3.2.5 Journal Usage Factor

Currently, the most well-known measure of journal quality is the Impact Factor from ISI. This figure is based on the citation data of a journal. Another measure, based on the same basic principles is the Hirsch Index. However, the Hirsch index is a measure of the impact of individual authors, based on their citation data (Shepherd, 2007a: 32). A new proposal is that a measure based on usage data, instead of citation data, would be an alternative way to evaluate the quality of journals (Sack, 2003/2004: 40; Shepherd, 2007b: 75).

To this end, COUNTER has been working with the UKSG (United Kingdom Serials Group) on developing the Journal Usage Factor. This is a new measure which would be based on COUNTER usage data from *Journal Report 1 (JR1): Number of successful full-text article requests by month and journal* and calculated according to the following formula:

$$\textbf{Usage Factor} = \frac{\textbf{Total usage (COUNTER JR1 for a specified period)}}{\textbf{Total number of articles published online (during a specified period)}}$$

A final report of the investigation was released in May 2007 and is available online at http://www.uksg.org/sites/uksg.org/files/FinalReportUsageFactorProject.pdf [2008, January 30]. The report finds that there is interest in developing a measure for journal quality (Journal Usage Factor) and author impact (Individual author Usage Factor) based on usage data instead of citation data (Shepherd, 2007a: 32).

3.2.6 Other initiatives

There are other electronic resource measurement projects that have been conducted that have not been covered in this chapter. Some of them are listed here along with their web sites.

- ARL New Measures Initiative
  http://www.arl.org/stats/initiatives/ [2008, February 6]

- NISO Forum on Performance Measures and Statistics for Libraries and NISO Standard Z39.7
  http://www.niso.org [2008, February 6]

- EQUINOX - Library Performance Measurement and Quality Management System
  http://equinox.dcu.ie/ [2008, February 6]

3.3 Summary of Chapter Three

Chapter Three discussed two of the main initiatives that have taken place in order to bring some sort of standardisation to the field of electronic resources usage statistics. The first of these was the ICOLC Guidelines for statistical measures of usage of web-based information resources. It is now the COUNTER Codes of Practice that are setting the standard for usage requirements.

Other projects discussed in this chapter are ScholarlyStats, SUSHI and Journal Usage Factor. The following chapter will seek answers to the research questions from an international perspective through a literature review.

Chapter Four: Electronic resource statistics in the literature

4.1 Introduction

> *"... [W]hat is the point of analysing e-resource usage statistics? What is the purpose? Often this second question is raised with a tone of frustration or exasperation (Oh, what's the use?)"*
>
> (Peters, 2002: 39).

The above sentiment might be voiced by library staff involved in the collection of electronic resource usage statistics, for whom the time consuming processes (see page 31) seem to bear little fruit as they do not see the data being used (Blake & Schleper, 2004: 460-461).

This chapter seeks to find answers to the following questions in the literature:

Why do libraries keep statistics for electronic resources?

Which statistics are they keeping?

What are the issues and concerns regarding statistics for electronic resources?

4.2 Why do libraries keep statistics for electronic resources?

Shepherd (director of the COUNTER project) provided a useful outline of the reasons why librarians need usage statistics:

1) To "assess the value of different online products/services";
2) to "make better-informed purchasing decisions";
3) to "plan infrastructure and allocation of resources"; and
4) to "support internal marketing and promotion of library services" (Shepherd, 2006: 142).

He also listed reasons why vendors need and develop usage statistics. These enable them to:

1) "experiment with new pricing models that reflect the current value of online publications, rather than the historical hard-copy holdings from which they were derived";
2) "assess the relative importance of the different channels by which information reaches the market";
3) "provide editorial support for new product development, etc";
4) "plan infrastructure, improve site design and navigation";
5) "obtain improved market analysis and demographics" (Shepherd, 2006: 143).

Shepherd's reasons for vendors requiring usage statistics are included here in order to give a holistic view of the generally accepted reasons for having usage statistics. However, the writer will be focusing on statistics from the librarians' point of view. Further discussion in this section will be structured according to Shepherd's outline of why librarians need to keep statistics for electronic resources, as quoted in the first paragraph.

4.2.1 To assess the value of different online products/services

A way of assessing the value of a resource is to calculate the cost-per-use. This data can be used to determine the cost of information, by performing calculations based on usage data and subscription costs (Blake & Schleper, 2004: 463; Franklin, 2005: 241; Sack, 2003/2004: 40). For example, knowing the subscription cost of a resource and the number of downloaded full-text articles or number of searches conducted, enables the librarian to calculate the cost per full-text article downloaded and cost per search. This information could then be used to compare resources based on cost-per-use.

It is however necessary to distinguish between good and bad usage. What is a good price per download? What is a good number of searches per potential user per year (Moen, Oguz & McClure, 2004)? As discussed on page 33 it is essential that all use statistics are looked at in context. What appears to be a good price per download in one country might not look so good in another. In terms of cost-per-download, this might be compared to the cost of obtaining the average article on interlibrary loan.

It is important that the idea of the value of a resource should not become confused with the cost of the resource. In the print environment the real cost of a journal includes the subscription cost, issue processing costs and the costs of shelving. In the electronic environment the costs of issue processing and shelving are no longer a factor. Electronic journals have additional features that add to their value: linking can take place from journal to journal, article to article, indexing database to full-text article, keyword alerts, tables of contents are available via RSS feeds (Sack, 2003/2004: 40). All these facilities enhance their usability. Ferguson makes the point that we should not restrict the way of looking at electronic journal use to the way we looked at print journal use. In the electronic environment a "network of information" is created as articles are interlinked and cancellation of an electronic subscription could cause a break in the linking possibilities (2003: 33).

4.2.2 To make better-informed purchasing decisions

Usage data can be used to assist in decisions around purchasing of library material in the collection development process and to support collection management decisions. Cost-per-use data can also be used to assist with decisions regarding subscription renewals (Stubbings & Hamblin, 2004: 25). High usage allows libraries to demonstrate that money is being spent on the correct resources (Ashcroft and McIvor, 2001:381; Lichtenberg, 2004: 12; Wisniewski and Fichter, 2007:54). Low usage figures, or declining usage, could be used to identify resources, databases or individual electronic journals, that are no longer of primary importance to the institution (Renwick, 2005:24), either due to a change in curriculum and research focus, or because another resource is filling the

information need. Turnaways, or rejected sessions, can be used to determine when simultaneous user numbers need to be increased for subscriptions that follow this model of purchasing (Hiott, 1999:44; Taylor-Roe & Spencer, 2005: 124). The subscriptions to certain resources are priced according to the number of simultaneous, or concurrent, users that may access the resource at one time. Another subscription model allows unlimited numbers of users from the subscribing institution to access resources at one time.

Usage data can also be used to identify individual electronic journal titles that should be added to the library's collection. The OpenURL resolver SFX can produce a usage file that lists the titles that are not held in full-text by the library, but that occur most frequently in users' results lists. By adding these titles to their holdings, the library would be able to serve their users' needs more immediately as the user would then not be reliant on an interlibrary loan request to fulfill their information needs.

Even though it is very difficult to get accurate usage data for print journals, libraries have compared cost-per-use data for electronic journals to their estimated figures for print journal use in order to justify the move from print subscriptions to their electronic editions (Franklin, 2005: 241; Gallagher, Bauer & Dollar, 2005:177).

The search reports that are available from some resources make possible the analysis of the "number of keyword and subject searches, the top 10 keywords and subjects input by users" (Lingle, 2005:52). This information could help librarians by alerting them to the research that is being undertaken at the institution (Pesch, 2004b:146; Sack, 2003/2004: 38) so that they might develop their collections accordingly.

4.2.3 To plan infrastructure and allocation of resources

Statistics and performance indicators can be used to "inform and enhance day-to-day decision-making, service development, reporting, marketing and advocacy" (Barton, 2004: 138). Usage data can assist management in deciding whether to develop electronic services or not, and to determine the cost effectiveness of electronic resources (Barton,

2004: 138; Bertot, 2004: 31, 32). It is vital for "overall planning and decision-making" to look at electronic usage statistics (McClure, 2004: 166).

If a library does not count visits to its website as visits to the library, as well as counting all virtual reference interactions and use of electronic resources, it is not reporting all of its activities (McClure, 2004: 166). In order to report fully on the activities of hybrid libraries it is essential that data on the use of electronic resources be collected along with the traditional performance measures (Bertot, et al, 2004: 30). Many libraries may notice an apparent decline in library use as they move resources into the electronic environment. For this reason it is important to collect usage figures for electronic resources in order to demonstrate that the library is still fulfilling a service (Nisonger, 2000:302).

Davis has suggested that the number of full-text downloads from electronic journals can tell librarians how many users are using the electronic journals. This would assist librarians in "evaluating the utility and impact of their journals subscriptions" (Davis, 2004: 380) as they would then be able to estimate the number of individual users who would be affected by decisions regarding electronic journals; for example cancellations of electronic journal subscriptions (Davis, 2004:388). Davis's investigation showed that there is a correlation between the number of full-text downloads at an institution and the number of unique IP (internet protocol) addresses doing the downloading at that institution (Davis, 2004:389). An increase in the use of individual journal titles can be a signal to the library of new research fields at the institution. The library can then be proactive to plan services accordingly (Taylor-Roe & Spencer, 2005: 124).

4.2.4 To support internal marketing and promotion of library services

"[I]nfrastructure and marketing challenges" can be highlighted by usage statistics (Hiott, 1999:44). Libraries can use usage statistics to greater advantage in the process of marketing their services (Helinsky, 2007:110). Low usage would tend to indicate a resource that needs to be marketed (Stubbings & Hamblin, 2004: 25). Lack of use could

be either an indicator that the resource is not required by the users, or it indicates that users are not aware of its availability (Pesch, 2004b: 145; Renwick, 2005:24).

Usage data also allow libraries to quantify their services and to demonstrate this to stakeholders (Tenopir, 2003: 33). At the University of Hong Kong, usage statistics "enable the libraries to tell the faculty what ... [they] are getting for ... [their] dollars and ... inform the staff which databases ...[are] being used and which ...[are] not" (Ferguson, 2003: 28).

Statistics can be used by a library "to describe and represent its activities in the networked environment" (Bertot, McClure & Ryan, 2001: 75). Bertot, McClure and Ryan make the point that it is important to identify a set of statistics for the networked environment that libraries can use for the purposes of benchmarking their services against those of similar organisations (2001: 2).

4.3 Which statistics are libraries keeping for electronic resources?

In an article written in 1999, and therefore quite dated in terms of developments around electronic resource usage statistics, Hiott discusses usage statistics at the Houston Public Library. Statistics being kept at the time were: sessions, logins, searches, documents downloaded, and turnaways. (Hiott, 1999:44). Over the years these are still the core statistics that librarians are keeping, and will be discussed in greater detail below.

4.3.1 Sessions

COUNTER defines a Session as follows

> *A successful request of an online service. It is one cycle of user activities that typically starts when a user connects to the service or database and ends by terminating activity that is either explicit (by leaving the service through exit or logout) or implicit (timeout due to user inactivity)*
>
> (COUNTER, 2005b)

The ICOLC Guidelines imply that the terms *sessions* and *logins* can be used interchangeably (Blecic, Fiscella & Wiberley, 2007: 27; ICOLC, 2006). However, it is clear from the COUNTER definition that a session is far more that just logging into a database, and usually involves some activity once logged in.

Cost-per-session figures can be calculated by dividing the subscription cost by the number of sessions. This information allows the librarian to put a measure of value to a resource (Conyers, 2004: 151; Moen, Oguz & McClure, 2004).

Commercially produced alphabetic lists of libraries' electronic journal holdings are another source of usage statistics. EBSCO's A-to-Z administration system is able to produce session, search and link-out reports that can be run according to various time frames. Session reports carry information on the number of sessions opened, including details of "average session length, total number of pages viewed, and average number of pages viewed per session." LinkOut reports record the number of times links have been followed (Lingle, 2005:52).

4.3.2 Searches

The COUNTER definition of a Search is

> *A specific intellectual query, typically equated to submitting the search form of the online service to the server*
>
> (COUNTER, 2005b)

The number of searches conducted can indicate the level of use being made of the resource and can be used to compare the importance of similar resources. Pesch suggests that the number of searches conducted could be equated to reference questions answered by librarians (2004b: 146).

Cost-per-search figures can be calculated by dividing the subscription cost by the number of searches conducted. This information allows the librarian to put a measure of value to a resource (Conyers, 2004: 151; Moen, Oguz & McClure, 2004).

4.3.3 Documents downloaded

The COUNTER term for *document* is *Full-text article*. The definition is as follows

> *The complete text, including all references, figures and tables, of an article, plus links to any supplementary material published with it*
>
> (COUNTER, 2005b)

The number of full-text articles can be further broken down according to the way the articles are formatted. COUNTER lists the following options: HTML, PDF, Postscript, References (COUNTER, 2005b).

The number of full-text articles downloaded is used by libraries to demonstrate the level of use of resources. Plotting the number of full-text downloads over time can illustrate trends of use in the library. Combining the number of full-text articles with the subscription cost can provide information on the cost per download and this can be used to measure the value of the information accessed (Bevan, Dalton & Conyers, 2005: 117-121; Moen, Oguz & McClure, 2004).

4.3.4 Turnaways

> *A turnaway (rejected session) is defined as an unsuccessful log-in to an electronic service due to exceeding the simultaneous user limit allowed by the license.*
>
> (COUNTER, 2005b)

Turnaways (Bertot, et al, 2004: 31) can be used to identify resources that need to have their subscriptions upgraded to allow for an increase in the number of simultaneous users for those resources (Pesch, 2004b:145). Libraries might begin their subscription to a resource with a small number of simultaneous users in order to gauge the level of real interest at the institution. A high number of turnaways indicates a real need for access to the resource and the subscribing library should consider spending more on their subscription in order to increase the number of simultaneous users.

4.3.5 Use from within the library versus access from elsewhere

Some libraries are interested in reporting usage of electronic resources by location. This might be usage by branch library (Hiott, 1999: 45), or usage by remote access. Electronic usage reported by branch library would enable the branch librarians to identify which resources might be in need of promotion at the branch, if usage is low. Usage reported by branch also allows the branch librarians to study the pattern of use for electronic resources at the branch and will allow them to plan their services accordingly.

4.3.6 Number of electronic resources

Information on the numbers of electronic library holdings in the various formats allows for the benchmarking of services and collections (Conyers, 2004: 149; Pesch, 2004b: 144). Counting resources by subject can identify areas that need collection development and help with collection sharing. Knowing the numbers of journals in various collections or databases can allow the librarian to perform overlap analysis to ensure that there is a minimum of duplication in purchased resources (Pesch, 2004b: 144).

4.3.7 Expenditure on electronic resources

Expenditure on electronic resources, broken down according to databases, electronic journals and electronic books (Bertot, et al, 2004: 31) can allow for benchmarking against peer institutions.

4.3.8 Virtual visits

A "virtual visit" is a "hit on the library website" (Conyers, 2004: 152). These need to be counted and added to actual visits counted by the physical library (Bertot, et al., 2004: 32) to give a true reflection of the use being made of the library services.

4.3.9 Conclusion

The statistics discussed in section 4.3 are the main statistics on electronic resources that the author found libraries in the United Kingdom and USA to be keeping. Chapter Six includes a discussion on whether these or any other statistics were also being kept by libraries in South Africa.

4.4 What issues and concerns are there with regard to statistics for electronic resources?

The following issues and concerns with regard to statistics for electronic resources have been identified from a survey of the international literature.

4.4.1 Lack of standardisation

The lack of standardisation in terms of electronic use statistics was an issue back in 1999 (Hiott, 1999: 47) and is only gradually being solved by the growing adoption of COUNTER compliancy by publishers. Standardisation is not yet fully realized, with

some publishers/vendors being COUNTER compliant and others not (Ferguson, 2003: 32; Schmidt, 2006). When there is such a lack of uniformity, comparisons between resources cannot be undertaken (Bauer, 2001: 37; Franco, 2005).

There are also problems with interpreting the usage data. This can be due to the way the information is collected and the lack of standardised terminology in the area (Franco, 2005). The basic terms need to be clearly defined (Peters, 2002: 44). The COUNTER Codes of Practice contain precise definitions of all applicable terms and their guidelines on how publishers should collect and report usage data ensure that standardisation in this area is increasing.

4.4.2 A time-consuming process

The collection of usage statistics is a time-consuming process (Ferguson, 2003:32) as it typically involves going to the publisher's website, logging in to the administration system, locating the files of relevant data, downloading them to a local computer and collating them into a programme such as Microsoft Excel. Back in 1999, Hiott commented that the collection of use statistics is a labour intensive process and suggested that the "next objective (after the data collection and performance measures are standardized) should be the automation and standardization of the reporting activities." (1999:47) ScholarlyStats and SUSHI are meeting this objective.

Because it is such a time consuming process the question has to be asked "[I]s the value of the measure's results worth the effort necessary to obtain them" (Bertot, McClure & Ryan, 2001: 51)? Librarians want to be assured that what they are counting is worthwhile (Conyers, 2004: 149) and that their statistics may be put to meaningful use.

4.4.3 Reliability of data

Drawbacks in terms of electronic usage statistics are that there is usually at least a month's delay in them becoming available. Data can also be corrupted or missing (Blake & Schleper, 2004: 461; Ferguson, 2003:30). These issues affect the confidence that librarians are willing to place in the reported usage data. An e-mail to the ProQuest-CSA mailing list in December 2007 is a case in point:

> We discovered some omissions of usage data that was collected during our recent peak months. We are reloading all usage data from September 2007 onward. Customer's usage reports may or may not be impacted as a result of the reload
>
> (ProQuest, 2007).

Concern has been raised whether publishers can be trusted not to inflate usage reports in order to ensure continued subscriptions (Schmidt, 2006). The independent auditing that is undertaken as part of COUNTER compliancy should address this concern.

There is also concern about how metasearching might impact on the usage statistics for individual electronic resources (Schmidt, 2006). Metasearching, or federated searching, is the ability to search across a range of resources that need not be on the same platform (Emery, 2005: 139). The numbers of searches recorded for resources that are metasearchable could be inflated by users who do not discriminate between databases when selecting which to cross-search (Stubbings & Hamblin, 2004: 29). Metasearch systems (for example MetaLib from Ex Libris or Searcher Analyser from TDNet) present users with a list of all the databases they might choose to search. Users are then able to pick and choose from the list, or they can *Select All*, as they decide on which databases they would like to search. The user enters a search request on the metasearch system which then performs searches on all the selected databases. The search results from the selected databases are then made accessible to the user from the metasearch engine interface.

4.4.4 Statistics must be viewed in context

"...we need to be careful about any inferences we make from an analysis of usage data about the needs, interests, and preferences of users" (Peters, 2002: 44).

There is great concern that libraries might make cancellation decisions based purely on cost and usage statistics. The caution is that the usage data must be looked at in context (Franklin, 2005: 245; Kraemer, 2006: 164; Schmidt, 2006). For example "one should not focus on raw numbers, but use compared to similar journals" (Nisonger, 2000: 300).

The "access infrastructure" (off campus access; OpenURL resolvers, metasearch systems) has an effect on usage and those journals with sophisticated linking functionality tend to show higher use (Kraemer, 2006: 169). The prominence a resource is given on a library webpage will influence how easily it is seen by users, thereby increasing its usage (Ferguson, 2003: 33). Sometimes a library will continue to subscribe to an expensive resource with high cost-per-use figures because it is an important resource for a specific department (Franklin, 2005: 245; Stubbings & Hamblin, 2004: 25).

Peters has suggested a very neat guide to the contextualisation of usage statistics and which assists in showing how statistics in specific contexts become more meaningful. The following are his "basic contexts into which these usage reports can be placed:

> *Resource context.* How is usage of the e-resource evolving over time? ... [This facilitates the identification] of long-term trends ...
>
> *Temporal context.* In order to successfully understand, interpret, and apply an e-resource usage report, the analyser must have a thorough knowledge of the e-resource content, interface, search engine, and general structure.
>
> *Similar resource context.* Usage statistics for similar e-resources can be compared and contrasted ...
>
> *Peer institution context.* Usage statistics for the same e-resource from peer institutions can be compared.

Print counterpart context. Sometimes the usage statistics for the print counterpart to an e-resource can be compared with the e-resource usage statistics. There are inherent dangers and limitations with this method of contextualization, however" (2002:44).

The above contexts provide a very useful guideline for thinking about ways in which to analyse usage statistics. The temporal context is a difficult area to manage. Issues that fall into this area include, amongst numerous others, keeping track of:

- when large amounts of content are added to a database;
- when back-files (earlier years) are added to electronic journal access;
- when access to a resource is cut off for any reason for any length of time; and
- when publishers change their systems and ways of counting usage (Ferguson, 2003: 33).

Librarians have to keep a record of all the changes that occur in electronic resource collections so that this information can be viewed alongside the usage statistics and provide the context required for meaningful analysis.

4.4.5 Management of statistics

Many writers are of the opinion that if used appropriately, electronic resource usage statistics could provide valuable data to libraries. However, libraries need some sort of assistance in order to do manage these statistics efficiently. Many of the Electronic Resource Management (ERMs) systems already on the market or in development include the capacity to store the usage figures and related information (Emery, 2005: 141; Medeiros, 2005: 146).

4.4.6 How to count electronic resources holdings

Although it would seem a simple thing to be able to report the number of electronic resources a library holds, there are complications. When it comes to counting databases, what constitutes a database? Does a collection of full-text journals constitute a database?

It is often difficult to keep track of the journal titles and numbers of journal titles included in big deals, publisher collections and aggregated databases. The journal titles in these collections count towards the number of electronic journals held by the library even though the library does not actually subscribe to the individual titles. Another complication is whether a library counts a print plus electronic subscription as a print journal subscription or as an electronic journal subscription, or as both. This question is particularly pertinent when calculating expenditure on electronic resources (Conyers, 2004:150). The items that are to be counted clearly need to be well defined.

4.5 Summary of Chapter Four

Since the advent of electronic resources, it has been generally recognised that it is possible to obtain usage statistics from the computer systems that run the resources. For a long while however, these statistics have not lived up to librarians' expectations that they would be easy to use, primarily due to a lack of standardisation. The purpose of this literature survey has been to identify which statistics for electronic resources libraries have been gathering and using, for what purposes they have employed the information gathered, and what problems they have encountered along the way.

Whereas Chapter Four has looked at the international situation through the literature, the next two chapters will address the situation in South Africa. Chapter Five is a discussion of the South African library environment, with specific reference to electronic resources in academic libraries.

Chapter Five: The South African academic library environment

5.1 Introduction

Chapter Five serves as an introduction to Chapter Six where a survey of the use of statistics on electronic resources in South African academic libraries is discussed, and aims to provide background information on the academic library environment in South Africa in order to provide a South African context for the research questions posed in Chapter One.

SANLiC, the consortium of academic libraries in South Africa is discussed. This chapter also includes a section on the draft *Measures for quality in SA HEI libraries* from the Committee for Higher Education Librarians in South Africa, CHELSA. The final section of the chapter presents a case study on the management of electronic resources usage data from the author's institution, the University of Cape Town.

5.2 SANLiC

The Coalition of South African Library Consortia (COSALC) was established in July 1999 (South African National Library and Information Consortium, [2007a]) in order to unify the efforts of the five South African higher education consortia that had been established between 1992 and 1998 (Thomas, 2007: 82, 83). The South African regional academic consortia and other stakeholders who make up the membership are: Cape Higher Education Consortium; Foundation of Tertiary Institutions of the Northern Metropolis; Eastern Seaboard Association of Tertiary Institutions; Eastern Cape Higher Education Association; Free State Higher and Further Education and Training Trust; South African National Research Information Consortium; National Library of South Africa; and, the Library and Information Association of South Africa. (South African

National Library and Information Consortium, [2007b]) A table of the institutions that constitute the different consortia is included on page 44.

During 2005 and 2006 COSALC underwent a restructuring process and in 2007 the body was renamed the South African National Library and Information Consortium, SANLiC (COSALC, 2007) to broaden the focus and to become more accessible to membership beyond academic and research consortia (Thomas, 2007:83).

The main project of the former COSALC was SASLI, the South African Site Licensing Initiative. At its Annual General Meeting in November 2006 it was agreed, amongst other things, that the SASLI name would fall away to be replaced by SANLiC-SL (SANLiC Site Licensing) (COSALC, 2007). At the time when the questionnaire that will be discussed in Chapter Six was originally compiled and distributed, the name in use was SASLI.

The SASLI office was launched on 30 July 2002 with a full-time project coordinator (COSALC, 2002) who negotiated consortium prices and licenses for electronic resources on behalf of the members of the consortium. When the Coordinator resigned in the latter half of 2006 there were 73 items on the list of electronic resources on offer through SANLiC (SASLI, 2006). Since the resignation there has been little new activity out of the SANLiC office in terms of agreements for additional resources. An office administrator has been performing a maintenance role for existing product offerings. The post of SANLiC Manager was advertised in 2007, with a closing date of 30 April (South African National Library and Information Consortium, 2007c). Unfortunately, no-one was appointed and the position still lies vacant in January 2008. With no-one obviously coordinating activities, publishers and vendors are now approaching libraries directly. In the past they would have been referred to the SASLI office and libraries would then be assured that they were all being offered the same terms and conditions. In the current environment the librarians who deal with electronic resources are missing the oversight that the then SASLI Project Coordinator provided.

In 2006 SASLI was instrumental in organizing two learning opportunities for librarians on electronic resources usage statistics as a need for this had been expressed by the regional consortia. There was a session on statistics at the two-day seminar held at the CSIR Conference Centre, Pretoria, in July. The title of the event was *An integrated approach to e-content: structure through technologies and standards*. The programme included sessions on Project COUNTER and ScholarlyStats (see p.17). At around the same time Peter Shepherd (Director, Project COUNTER) visited the South African regional library consortia to in South African librarians about COUNTER. (COUNTER is discussed on page 13.)

5.3 CHELSA Measures for quality in South African Higher Education Institution libraries

The Committee for Higher Education Librarians in South Africa (CHELSA) is a body of library directors from the higher education institutions in South Africa. It was formed in 2004 (Thomas, 2007: 79).

The Quality Assurance Sub-committee of CHELSA was briefed to compile a document entitled *Measures for quality in SA HEI libraries*. It is based on measures from ARL (Association of Research Libraries), CAUL (Council of Australian University Librarians) and SCONUL (Society of College, National and University Libraries) (Committee for Higher Education Librarians of South Africa, 2006) and proposes a set of standardised measures for collecting statistics in South African Higher Education libraries. Draft three of the document is available at http://www.ched.uct.ac.za/cil/dils/resources.html [2007, December 20].

The measures proposed for electronic resources in this document are as follows, keeping the section numbering of the document:

3. SUBSCRIPTIONS TO ELECTRONIC AND CONTINUING RESOURCES

3.1 Number of serials received by subscription

3.1.1 Number of serials received as printed items only

3.1.2 Number of serials received in electronic form only

3.1.3 Number of serials received in both print and electronic form

3.2 Number of electronic databases received by subscription

3.3 [Number of] electronic books

9. USE OF LIBRARY SERVICES

9.6 Electronic Transactions

9.6.1 Number of successful requests for full-text articles

9.6.2 Number of successful accesses to electronic books (Committee for Higher Education Librarians of South Africa, 2006)

The document clearly defines what databases are and how to count subscriptions to electronic journals. In terms of electronic resource usage it measures what is probably the most important indicator of the value of a resource, the number of full-text articles.

However, not all electronic resources contain full-text content; there are databases that are strictly indexing and abstracting databases. The CHELSA measures only specify two types of electronic transactions (measures 9.6.1 and 9.6.2 above) that should be recorded. They both focus on accesses made to full-text content that is available in either electronic journals or electronic books. Indexing and abstracting databases can be important search and discovery tools for the full-text content that is being accessed and counted in measure 9.6 of the CHELSA document. Pesch has suggested that the number of searches conducted on electronic databases could be equated to reference questions answered by librarians (2004b: 146). By including the *number of searches* under the measures recorded for electronic transactions, a more inclusive view of resource usage would be given by the CHELSA measures.

5.4 Case study: Electronic resources statistics at UCT Libraries

This case study of electronic resource statistics at the University of Cape Town Libraries is included as an illustration of how one library is managing electronic resource statistics. It is from the author's experience in dealing with the issues of electronic resource usage

statistics that the research questions originated. The author wanted to learn how to improve the processes at her home institution.

The author has filled the role of Electronic Resources Librarian at the University of Cape Town since 2000. She was the first incumbent of this position at the institution and as such has been involved in the development of the processes regarding all aspects of electronic resources. With other, more urgent, administrative tasks that accompany the management of electronic resources, the collection of usage statistics was always lowest on the priority list of tasks that needed to be completed.

In 2002 a colleague was given the job of collecting and collating the usage statistics for all the databases held by UCT Libraries. Slow internet speeds in South Africa add to the conditions that make the collecting of usage data a laborious process. The usage data are gathered and saved into Microsoft Excel spreadsheets and made accessible to library staff on a common drive on the library network. Librarians are able to view the usage information themselves in order to see how the resources are being used, and to make decisions on the services and resources offered in their subject areas. For example, the librarians who have science as their subject responsibility are able to identify which of the databases with a science focus have low usage figures. They can then decide whether these databases should be promoted to their users, or suggest that the databases be cancelled.

Since 2000, at the regular meetings of the Electronic Resources Group within UCT Libraries, a standing item for discussion has been usage statistics. In 2005 the schedule of meetings of this group was revised and specific themes attached to meetings. The mid-year meeting of the group is now devoted to considering the usage statistics for the previous year. Usage data are presented to the meeting in the form of graphs and trends are discussed amongst the librarians present. The purpose of the meeting is to generate discussion around the use of electronic resources. Resources that show low use, or whose use has decreased, are looked at with a view to possible cancellation or earmarked for intensive promotion to the user community.

The procedures described above refer mainly to the usage data that are available for databases. Although statistics for electronic journals are available, these are not being routinely collected at UCT Libraries. In 2007 a subscription to ScholarlyStats was opened at the minimum subscription level of nine platforms. ScholarlyStats collects the monthly usage statistics from each of the nine platforms and collates the information on the ScholarlyStats web site. The platforms UCT chose for this service were electronic journal platforms that were not already being harvested for usage data by the UCT library staff. The idea was to start out small and test the service. If it was demonstrated to be worthwhile the subscription level could be increased to cover more platforms. The subscription to the service was renewed at the same level for a second year as it was felt that one year's worth of data was not sufficient for meaningful analysis of the service and the statistics.

5.5 Summary of Chapter Five

The South African academic library environment was introduced in Chapter Five with an introduction to the South African National Library and Information Consortium, SANLiC. Also included was a discussion of the *Measures for Quality* document compiled by the Quality Assurance Sub-committee of the Committee for Higher Education Librarians in South Africa. The last section in the chapter was a case study of electronic resource management at the University of Cape Town Libraries. Chapter Five has set the scene for the survey of South African Libraries that will be covered in the following chapter.

Chapter Six: Survey of South African academic institution libraries

6.1 Introduction

The survey of South African academic libraries is discussed in this chapter. The findings are reported and analysed with reference to the research questions that were posed in Chapter One.

6.2 Purpose of investigation

The reason for surveying South African academic libraries participating in SANLiC was to establish to what extent South African libraries are keeping and using the same statistics for electronic resources as their colleagues internationally. A questionnaire was designed based on the questions that the author addressed in Chapter Four of this dissertation. It was expected that the answers would be similar, as the issues librarians face regarding electronic resource statistics are similar worldwide. If the survey results made it apparent that there is a need amongst this group of institutions for training in any aspect of electronic resource statistics, the author would refer the matter to SANLiC for the coordination of training interventions.

6.3 Methodology

On 7 November 2006 a questionnaire (see Appendix) was e-mailed to the 23 South African academic institutions that form SANLiC, (see Table 3, page 44). A reminder e-mail was sent on 3 January 2007 to those institutions that had not responded to the original message. On 7 August 2007 the questionnaire was again sent out to non-respondents. In the end a total of 15 responses was received thus yielding a response rate of 65%. It is acknowledged that the number of responses is small, but as the instrument

was sent to the whole population, i.e. all the South African SANLiC institutions, and responses were received from across the range of institutions, it does, as Table 3 indicates, give a reasonably accurate picture of the South African academic library situation.

[Table 3 follows on the next page.]

Table 3: Academic libraries surveyed

Institution	Responded
Cape Peninsula University of Technology	yes
Central University of Technology, Free State	yes
Durban Institute of Technology	yes
Mangosuthu Technikon	yes
Nelson Mandela Metropolitan University	yes
North-West University	no
Rhodes University	yes
Tshwane University of Technology	yes
University of Cape Town	yes
University of Fort Hare	no
University of Johannesburg	yes
University of Kwa-Zulu Natal	yes
University of Limpopo	no
University of Pretoria	yes
University of South Africa	yes
University of Stellenbosch	yes
University of the Free State	yes
University of the Witwatersrand, Johannesburg	no
University of Venda for Science and Technology	no
University of Zululand	no
University of the Western Cape	yes
Vaal University of Technology	no
Walter Sisulu University for Technology and Science, Eastern Cape	no

6.4 Analysis of responses

Respondents were asked to give their names and institutions for administrative purposes and were also asked to indicate their position. Confidentiality was assured and no individual institution would be identifiable from the discussion of the findings. Each question from the survey will be addressed in the following paragraphs.

6.4.1 Statistics for databases

The reason for asking Question One was to ascertain whether institutions are recording the database statistics that COUNTER requires from database publishers. Questions 1(a) through 1(d) refer to data that are stipulated in the COUNTER code of practice for journals and databases and are reported in the COUNTER Database Reports 1, 2 and 3. (See discussion of COUNTER on page 13) Two institutions did not answer the detail of this question, other than to say they only kept statistics for what they regarded as the most important databases. The analysis of the question has therefore been done on responses from 13 institutions.

[Discussion of Question One continues on the following page.]

QUESTION 1

What publisher/vendor produced usage statistics for electronic databases does your institution collect and keep? Check all that apply.

a)	Number of searches	13 YES
b)	Number of sessions	8 YES; 5 NO
c)	Number of fulltext downloads	12 YES; 1 NO
d)	Number of turnaways	11 YES; 2 NO

e) Please list any other database usage statistics that you keep.

The following comments were noted:

"no. of hits and no. of visits where applicable – HeinOnline"

"no. of e-mails; pdfs vs HTMLs"

"It depends on the database type: Aggregators: Usage statistics per database in the package; Publishers' Databases: Usage per journal; Total usage statistics for Growth Rates in usage; Content statistics per subject area; Cost per search & Cost per Download. This is also worked out for overall database costs."

"We keep all the above statistics but use Number of searches only at this stage for our purposes."

Answers show that most institutions are recording all the COUNTER statistics for databases. In addition, a few institutions are recording statistics provided by database publishers that are not required for COUNTER compliancy. These are listed above as responses for question 1(e). The third response is included for the sake of completeness in recording responses from the survey. However, this information would be more correctly placed as an answer to question 6.

The issue of statistics from HeinOnline (a database of electronic legal full-text journals) was raised in a free-text response to question 1(e). HeinOnline does not provide COUNTER compliant statistics. It provides monthly figures of "hits" and "visits" which it defines as follows:

"Hits are basically any time a page is accessed, and some pages may count as more than one 'hit'. When first going to heinonline.org, there are several graphics on this page, which will count as multiple hits. Once you get into the HeinOnline collection, and start your research, there are a number of pages that are hit prior to actually getting to the article you wish to read (in other words, you first need to find the article, and depending on the way you choose to access the article, you could add several additional hits to the site.

Visits are classified as the time from when someone logs on until they log off from HeinOnline (or remain inactive for more than 30 minutes)"

(Hannon, 2005)

At the University of Cape Town this kind of non-standard data is dealt with as follows. "HeinOnline visits" are renamed Sessions/Logins, the nearest COUNTER equivalent terminology, and recorded in the spreadsheet as non-COUNTER compliant. Hits are not recorded.

Another of the usage statistics reported in question 1 (e) is number of PDF and number of HTML full-text downloads. Although the COUNTER reports for databases do not require publishers to distinguish between PDF full-text article downloads and full-text articles downloaded in HTML, this distinction is present in the COUNTER Journal Reports, specifically Journal Report 1. The databases that are basically large collections of full-text journals, sometimes report the PDF and HTML downloads at a database level as well as at the journal level.

"No. of e-mails" is not a figure that many databases provide. However, where it is provided it is interesting to record as it gives an extra indication of how users are interacting with the database.

In answering question 1(e) a further institution made the comment that although they record all the statistics they only make use of the figure for the number of searches.

6.4.2 Statistics for electronic journals

Question Two placed the spotlight on electronic journals, whereas Question One had focused on databases. Electronic journals could be journals that are accessed as part of a database or collection of electronic journals, or they could be individually accessed. In both these scenarios the statistics discussed in Question 2 deal with the individual journal titles. Two institutions did not answer the detail of this question therefore the analysis for this question is done on responses from 13 institutions.

QUESTION 2

What publisher/vendor produced usage statistics for electronic journals does your institution collect and keep? Check all that apply.

a)	Number of searches	11 YES; 2 NO
b)	Number of sessions	4 YES; 9 NO
c)	Number of fulltext downloads	11 YES; 2 NO
d)	Number of turnaways	5 YES; 8 NO

e) Please list any other electronic journal usage statistics that you keep.

The following comments were noted:

"Number of downloads per Journal. With ScienceDirect we also need statistics of the number of article downloads per subscribed journals vs articles in the freedom collection. This determines international pricing."

"no. of e-mails; pdfs vs HTMLs"

"The statistics we keep for e-journals usage comes from our A-Z list."

"total no. of searches per month; top 50 ejournals linked from A to Z list"

"price per article"

Responses show that fewer institutions record statistical usage data for electronic journals than databases. Usage is recorded according to what is available from the publisher platform. For the most part the availability of data matches the information available for databases, except that the figures are recorded per journal title.

One institution commented: "With ScienceDirect we also need statistics of the number of article downloads per subscribed journals [versus] articles in the freedom collection". This refers to the subscription agreement between SANLIC and Elsevier Science for access to titles on ScienceDirect. In terms of the agreement, institutions commit to maintaining the individual subscriptions held at the time of joining the consortium. This value is a factor in calculating the subscription cost of the ScienceDirect package for each participating institution. In addition to the titles for which individual subscriptions are held, the consortium package gives access to the so-called Freedom Collection. This Collection is available to academic institutions and gives access to "all non-subscribed Elsevier journal content at a significantly reduced rate." (http://info.sciencedirect.com/licensing/additional/freedom/) [2008, February 6]

"no. of e-mails; pdfs vs HTMLs" has been discussed under section 6.4.1.

Two institutions mentioned counting usage from their electronic journal lists. Most of the commercial providers of alphabetical lists of electronic journals provide usage statistics for the list. Although COUNTER specifies that the responsibility for providing usage data lies with the publisher in this situation (see p. 15) the usage information that is available through these providers (for example SerialsSolutions and EBSCO) can provide a picture of how library users are accessing the electronic journals.

In the response to Question 2(e) one institution mentioned "price per article" as a statistic that they recorded for electronic journals. This is not a usage statistic but is a calculated figured based on the usage statistics.

6.4.3 Statistics for electronic books

Question Three was asked in order to establish whether libraries are also keeping statistics for electronic books to the same extent that they are doing for databases and electronic journals. All 15 respondents answered Question Three.

QUESTION 3

Do you keep statistics for electronic books? 4 YES; 11 NO

If yes, what statistics do you keep?

The following comments were noted:

searches and fulltext

"We only report overall statistics and not statistics per book title [for electronic book packages]."

Only four respondents indicated that they record statistics for electronic books. The information kept is the number of searches and number of full-text retrievals. The small number of institutions keeping statistics for electronic books is indicative of the slow uptake of electronic books within academic institutions in South Africa. Although more institutions are showing an interest in electronic books, many institutions are moving into this arena with caution. A recently published paper on the cataloguing of electronic books in South Africa concluded that a substantial number of librarians professed a "lack of experience and confidence" in cataloguing electronic books (De Jager, 2007: 52). It is suggested that this inexperience might also be extended to a limited ability in collecting statistics for electronic books.

6.4.4 Other electronic resource statistics

Question Four addressed statistical data that were not usage data. These data include information on the numbers of the various types of electronic resources and expenditures. Answers to this question were provided by 13 institutions.

QUESTION 4

What other figures do you keep? Check all that apply from the following:

a)	Number of database subscriptions	12 YES; 1 NO
b)	Number of electronic journal subscriptions	11 YES; 2 NO
c)	Number of electronic books purchased	8 YES; 5 NO
d)	Expenditure on electronic resources	11 YES; 2 NO
e)	Do you keep any other statistics for your electronic resources? Please list them.	

The following comments were noted:

"per search" stats

"Content per subject area is very important"

"Usage (no of searches) by type of resources – divided by: Bibliographic, Fulltext Abstracting and indexing, Fulltext journal publishers, E-reference sources, Ebooks, Research support databases (JCR, SerialsSolutions)."

From ScholarlyStats: "Usage by platform; Highest use journals from all resources (50 highest journal titles); Consolidated list of journal usage from all databases"

"Expenditure per department and per vendor. Expenditure over a period of three years."

Most institutions have access to records of the number of electronic resources to which they subscribe, and they can be differentiated by format, namely database, electronic journal and electronic book. Most institutions can also access data on the expenditure figures for these resources.

The free-text responses to Question 1(e) speak to dividing usage data into sub-divisions according to subject and type of resource. An institution also reported breaking down expenditure by department, which could be equated to subject, and by vendor or supplier. Statistical breakdown by subject gives the librarian an indication of how electronic resources are being used across the various subject disciplines. A breakdown by type of resource indicates how users are searching for information.

6.4.5 Staffing

Question 5 addressed the issue of what staff are involved in the process of collecting usage statistics for electronic resources. Twelve institutions provided answers to Question Five.

QUESTION 5

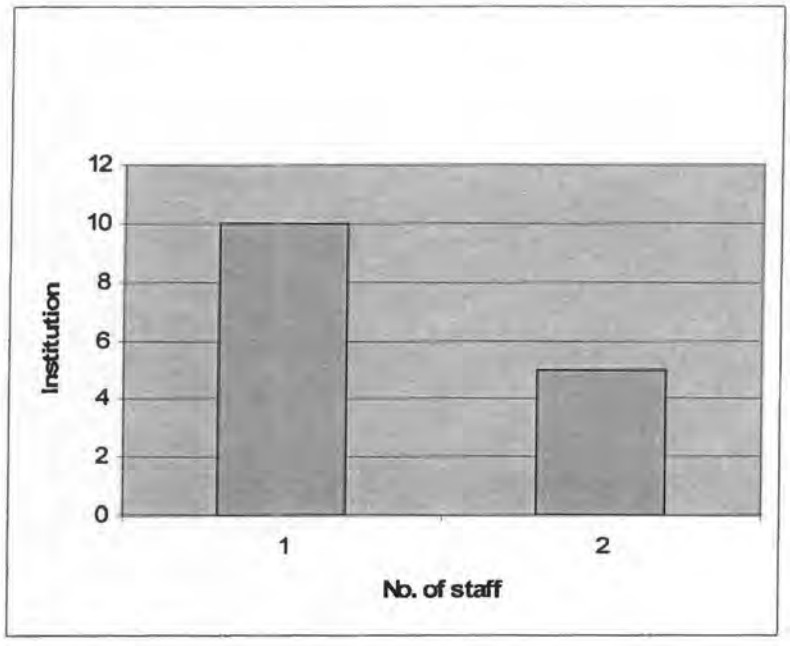
In the process of collecting your electronic resource statistics ...

a)	How many staff are involved?	Responses indicated that one or two staff were involved; see Figure 1 below.
b)	What level of staff are they?	They were professional librarians or para-professional staff.
c)	How much time do they spend collecting and managing statistics for electronic resources?	See figure 2 below.
d)	How are the usage statistics recorded and managed? Please explain.	See discussion below.

As illustrated in Figure 1, most institutions have one member of staff involved in collecting electronic resources statistics. A few institutions have two staff involved. Two institutions reported that para-professional staff do the collating of the statistics with the librarian doing the analysis and interpretation. One institution reported that their IT

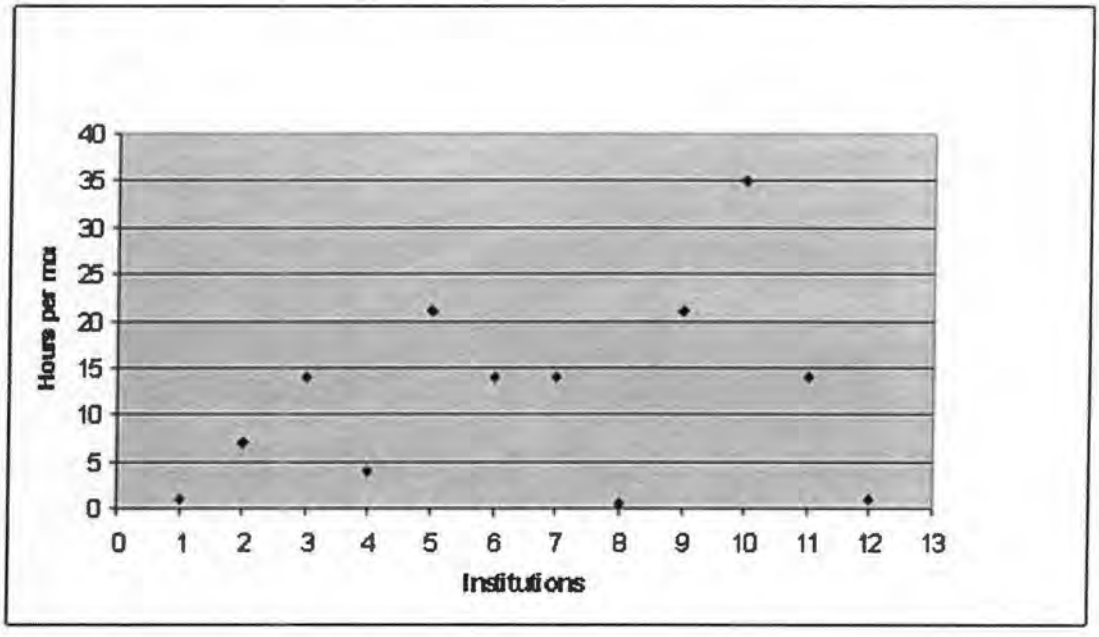
personnel deal with usage statistics. At 12 institutions it is the professional staff who attend to the whole task.

Figure 1: Number of staff collecting statistics



Responses to Question 5(c) were given in minutes, hours or working days. In order to represent the answers to question 5(c) in Figure 2, one day was converted to a working day of seven hours. Libraries reported spending between 30 minutes and 5 days per month on collecting and collating usage statistics. As Figure 2 illustrates most of the institutions spend approximately 14 hours per month collecting statistics.

Figure 2: Time spent collecting statistics



The answers to Question 5(d) indicate that some libraries have a very basic method of recording and managing usage statistics, while others have well-developed procedures. One institution did not answer question 5 (d) at all, and two gave answers that were unusable. Of the twelve institutions that gave meaningful answers, ten indicated that they used a spreadsheet to manage their statistics. Two institutions reported that they printed out the statistics from the vendor sites and filed them.

When using a spreadsheet institutions are recording the statistics by month. Four institutions reported that the statistics are made easily accessible to library colleagues on intranets or common network drives. Three institutions reported that they distribute statistics to key stakeholders in the institution.

6.4.6 Use of statistics for electronic resources

Question Six focused on how libraries were using electronic resource statistics. Although all 15 institutions indicated that they used the statistics for some purpose, two did not answer questions 6 (a) through (e).

QUESTION 6		
Do you use statistics on electronic resources? Yes or No.		15 YES
If YES, do you use statistics to:		
a)	demonstrate level of use?	11 YES; 2 NO
b)	demonstrate costs, eg. cost per fulltext download, cost per session, cost per search?	12 YES; 1 NO
c)	determine correct subscription level in terms of number of simultaneous users?	10 YES; 3 NO
d)	monitor number of turnaways?	10 YES; 3 NO
e)	For what other purposes do you use electronic resource statistics? Please explain.	
	The following comments were noted:	
to make decisions regarding renewals/cancellations: 3 institutions		
to compare databases: 1 institution		
to identify training and marketing requirements: 2 institutions		
to compare usage across institutions: 1 institution		

The literature review in Chapter Four has shown that low usage statistics or decreasing usage statistics can be used as indicators that training and marketing are required for certain resources. These same low usage figures might also be used as a reason to not renew subscriptions, although when used for this purpose the statistics must be looked at in context and not in isolation.

One of the responses to Question 6(e) was that statistics were used to compare usage across institutions. A few resources, especially those that are bought through the

Consortium allow a certain amount of comparison between peer institutions. Resources that fall into this category have been JSTOR and MathSciNet for the SANLIC group.

6.4.7 COUNTER awareness

Question Seven was included in the survey in order to establish the awareness of COUNTER amongst librarians. For an in-depth discussion of COUNTER please refer to page 13.

QUESTION 7
Are you aware of the Project COUNTER Initiative? How important is it for your purposes that usage statistics are COUNTER compliant? Please explain.

Out of 15 institutions, two responded to Question Seven that they were not aware of the COUNTER initiative. Although they had heard of the initiative, another two institutions did not know anything about it. The 11 institutions that knew about Project COUNTER counted it important as it imposed standardisation on the usage statistics produced by publishers.

Given the fact that Peter Shepherd (Director, COUNTER) did a series of presentations during July 2006 at the various centers throughout South Africa and at the SASLI seminar in Pretoria (*An integrated approach to e-content: structures through technologies and standards*, 26-27 July 2006), it is disappointing that not all institutions and librarians are fully versed in Project COUNTER and the advantages it brings to the world of electronic resource usage statistics.

6.4.8 Other issues

Question Eight, the last on the questionnaire, was an open question that allowed respondents to raise any issues that had not already been raised through the previous seven questions.

QUESTION 8
What other considerations/issues would you like to raise with regards to electronic resources usage statistics?

The following issues were raised by respondents:

- One librarian commented that there are still publishers who cannot provide libraries with usage statistics. This is particularly relevant in the South African situation with major publishers LexisNexis and Jutastat unable to supply this information. Although this issue was raised in one survey response, many South African Academic libraries have to contend with this situation as these are the major publishers of law information in the country.
- Libraries are considering developing their statistics processes. One institution reports looking into subscribing to ScholarlyStats to help manage their statistics. Two respondents already subscribe to this service. One institution is looking for ways to determine usage by faculty so that they will be able to target specific areas that need training or for the marketing of resources.
- "For the effort entailed in collecting stats, is it worth it to collect stats for all resources, or should the effort just be focused on a selection of key subscriptions?" This question was raised, or implied, by three respondents. Although collecting limited statistics might sound like a way to manage the amount of time spent on collecting usage statistics, it does mean that the library will be unable to accurately report the level of use being made of their electronic resources.

- Although huge strides have been made in terms of usage statistics for online resources, this information for networked CD-ROMs is lacking. This concern was raised by two respondents. As discussed on page 9, libraries resort to keeping manual statistics for CD-ROM databases, be they stand-alone or networked. It is highly probable that this manual count does not give an accurate portrayal of the use of the databases. These usage figures are also not comparable against usage figures for other resources.
- When libraries are short staffed, collecting usage statistics becomes very low priority. The institution that raised this issue was one that had recently undergone a merger and had not yet fully reassigned tasks appropriately. However, when it appears that little use is being made of usage data, busy staff everywhere will tend to place the task of collecting the information, low on their priority list.
- Some publishers have very user unfriendly systems for their statistics. The respondent who raised this issue cited the following examples:

  (1) StatsnetBase (CRC Press) is not easy to login to and the titles of reports are not descriptive;

  (2) A few publishers (AIP, IOP, RSC) are using CoreMetrics as their statistics platform. Apparently, the platform does not have a logout button. The user is prevented from accessing information on different publisher login codes for several hours, until their session is automatically timed out by the system.
- A further issue raised by the researcher relates to the way in which the survey questionnaire was completed. A couple of institutions completed the survey in a very rudimentary fashion. While this might be indicative of an unwillingness to complete survey questionnaires, or a sign of not having the time to do so, it could also suggest that the respondents did not have a good understanding of the collection and use of statistics for electronic resources. This could be an expression of a training need in this area, which is corroborated by De Jager's findings regarding the cataloguing of electronic books in South Africa. (2007: 53).

6.5 Summary of Chapter Six

Chapter Six reported on the findings of the survey that was conducted of South African academic libraries in order to answer the research questions from the South African point of view.

The reasons why libraries keep statistics for electronic resources and the ways in which they use this information tends to be the same all over the world. The concerns are also very similar. However, an additional concern raised in South Africa, is that there is a lack of knowledge amongst some librarians about some basic concepts to do with electronic resource usage statistics. The author suggests that training sessions should be organised to address this concern.

Chapter Seven, the final chapter of this dissertation, will be a wrap-up of the discussion that has taken place in Chapters One through Six.

Chapter Seven: Summary and recommendations

7.1 Introduction

Answers to the questions posed in Chapter One have been sought through the discussion in Chapters Two through Six. This final chapter summarises the findings of the preceding chapters and concludes with some suggestions on the way forward.

7.2 Summary of findings

The research questions posed in Chapter One are as follows:

Why do librarians keep statistics for electronic resources?

Which statistics are libraries keeping for electronic resources?

What are the issues and concerns with regards to statistics for electronic resources?

Through the literature review of Chapter Four, the author has sought answers to these questions from the international perspective. The e-mail survey discussed in Chapter Six addressed the questions from a South African point of view.

7.2.1 International findings

Shepherd's summary of why librarians should keep statistics for electronic resources provided a useful categorisation structure. All the reasons encountered in the literature review can be grouped together under his four reasons. These read as follows and were used to frame the structure of the discussion in the Chapter Four. Statistics for electronic resources are kept:

1) to "assess the value of different online products/services";
2) to "make better-informed purchasing decisions";

3) to "plan infrastructure and allocation of resources"; and

4) to "support internal marketing and promotion of library services" (Shepherd, 2006: 142).

The literature review identified the following statistics that libraries, internationally, are keeping for electronic resources:

- Sessions
- Searches
- Documents downloaded
- Turnaways
- Use from within the library versus elsewhere
- Number of electronic resources
- Expenditure on electronic resources
- Virtual visits

The first four data elements in the above list are those required by COUNTER. The rest of the items are not required by COUNTER, but are of interest to some libraries.

The issues and concerns about electronic resource statistics raised through the literature review can be summarised as follows:

- Lack of standardization with regard to the way usage is reported by publishers and the fact that not all resources are able to produce usage data;
- It is a time-consuming process to gather usage statistics;
- There are questions about the reliability of data;
- Caution is expressed that usage statistics should not be viewed in isolation, but must be looked at in context;
- Librarians need assistance with the management of usage statistics.
- There are intricacies in the counting process of numbers of electronic resources. The different items need to be clearly defined.

7.2.2 South African findings

As is evident from the answers received in response to the questionnaire as discussed in Chapter Six, many of the libraries of the academic institutions in South Africa conform to the rest of the world in terms of the practices of electronic resource usage statistics. The questions that are being asked by a number of South African librarians, with regard to the use of usage statistics for electronic resources are not very different from those encountered in the literature review discussed in Chapter Four.

7.2.2.1 Why do South African libraries keep statistics for electronic resources?

The reasons why a number of South African libraries keep statistics for electronic resources fall fairly neatly into Shepherd's outline of reasons much as those of international libraries do (Shepherd, 2006: 142).

a) To assess the value of different online products/services

As discussed on page 22 from the international perspective, the majority of South African libraries that responded to the survey report using usage statistics and cost of subscription to calculate cost-per-use figures, for example cost per search. These figures give the libraries a measure of the relative values of resources. This allows a certain degree of comparison to be made of resources.

b) To make better informed purchasing decisions

Eleven of the South African libraries that responded to the survey are using usage statistics to monitor the level of use being made of resources. Usage figures are being used to support renewal and cancellation decisions for subscriptions. Turnaway counts

are being used to ensure that subscription levels are at optimum for the resources that have a simultaneous user limit.

c) To plan infrastructure and allocation of resources

The author could not identify planning and allocation of resources as a reason for keeping statistics from the questionnaire responses from South African libraries.

d) To support internal marketing and promotion of library services

As reported in the questionnaire responses, two South African libraries use low usage figures to identify resources that require promotion to users. The circulation of usage reports to key stakeholders in institutions accomplishes the degree of marketing of library services. Where available, some South African libraries report that they are using usage statistics for specific resources to make comparisons with their peer institutions.

7.2.2.2 Which statistics for electronic resources are South African libraries keeping?

Like their international counterparts, a number of South African libraries report that they are keeping the following statistics for electronic resources:

- Searches
- Sessions
- Full-text downloads, including the distinction between PDF and HTML where provided
- Turnaways
- Number of subscriptions
- Expenditure
- Other statistics provided by publishers that are not COUNTER compliant: hits; visits; e-mails sent from database

No South African libraries reported keeping statistics on virtual visits.

7.2.2.3 What are the issues and concerns that South African librarians have regarding statistics for electronic resources?

The issues and concerns raised by South African librarians are similar to those discussed from page 30.

A concern that the author did not encounter in the literature review was concern over a lack of knowledge by international librarians. However, in South Africa, based on the responses to Question Seven, at some institutions there is still a lack of knowledge regarding one of the fundamental initiatives to do with usage statistics, COUNTER. There is obviously the need for training around this issue. Training would also be beneficial on how to make use of the statistics that are available. Seminars that address the theoretical issues, but which include practical aspects would be the most effective means of training.

One example of a suitable training programme is the *Usage Statistics Training Seminar*. This event, run by the United Kingdom Serials Group (UKSG) could be appropriate in the South African situation. After a session on COUNTER by the Project Director and one from the publisher's perspective, the remaining sessions are practical and hands-on. Led by librarians who share their experiences, they educate attendees on using usage statistics as a practical tool. The final session of the day-long seminar, "What will you do now? Outcomes to take back to base", wraps up the programme by summarising the issues (Sadler, 2007).

If librarians were more generally aware of how the usage figures for electronic resources could be used to benchmark their libraries against peers both nationally and world-wide, they might think differently about the amount of work involved with collecting the

information. They would then not be prompted to suggest that only statistics for select resources be kept, as discussed on page 57, under section 6.4.8.

Generally, it seems that institutions are recording the usage statistics that are required by COUNTER. With coordination at the Consortium level, it might be possible to get agreement from all participants on requirements for reporting usage data, thereby setting up the framework for establishing a benchmark for electronic resources in South Africa.

7.3 Recommendations

As an outcome of the research done for this dissertation, the following suggestions for future developments may be made:

7.3.1 For the University of Cape Town

In terms of the collecting of usage data, the author intends to put in place procedures that will allow for the collecting of all the usage data that is available for databases, electronic journals and electronic books. Having this information readily available will allow for the benchmarking of the UCT Libraries' services against institutions internationally who follow the data collection models of SCONUL and the ARL.

7.3.2 Nationally

It is apparent from the varied responses to the questionnaire discussed in Chapter Three that there are gaps in some librarians' knowledge regarding the potential uses for statistical data for electronic resources. Not all librarians are aware of Project COUNTER and the work being done to standardize electronic resource usage data. Ideas on how usage data might be used to promote the library's standing in the institution also need to be disseminated more widely. One way to do this is to be able to demonstrate how well

one's own library services are performing against peer institutions in the country. As discussed on page 38, the adoption of the CHELSA Measures for Quality in SA HEI libraries would facilitate this comparison. Once an agreement to this effect is implemented and being used for benchmarking across the country, it is suggested that training events could be scheduled under the auspices of SANLiC to ensure that the library staff who are responsible for collecting electronic resource usage data, understand the concepts and the greater context of what they are doing. Although usage data can be used to demonstrate trends with individual resources and to identify little used resources for cancellation or marketing, the author contends that the real power behind these figures lies in being able to compare library performance against peer institutions in a national and international arena. For the effort associated with collecting and collating usage data, the library staff doing the job should be made aware of how the information is, or could, be used to promote their library.

Bibliography

Ashcroft, L. & McIvor, S. 2001. Electronic journals: managing and educating for a changing culture in academic libraries. *Online information review*, 25(6): 378-387.

Barton, J. 2004. Measurement, management and the digital library. *Library review*, 53(3): 138-141.

Bauer, K. 2001. Indexes as tools for measuring usage of print and electronic resources. *College & Research Libraries*, 62(1):36-42.

Bertot, J.C. et al. 2004. Capture usage with e-metrics. *Library journal*, 129(8): 30-32.

Bertot, J.C., McClure, C.R. & Ryan, J. 2001. *Statistics and performance measures for public library networked services*. Chicago: American Library Association.

Bevan, S., Dalton, P. & Conyers, A. 2005. How usage statistics can inform national negotiations and strategies. *Serials*, 18(2): 116-123.

Blake, J.C. & Schleper, S.P. 2004. From data to decisions: using surveys and statistics to make collection management decisions. *Library collections, acquisitions & technical services*, 28: 460-464.

Blecic, D.D., Fiscella, J.B. & Wiberley, S.E. 2007. Measurement of use of electronic resources: advances in use statistics and innovations in resource functionality. *College & research libraries*, 68(1): 26-44.

Blecic, D.D., Fiscella, J.B. & Wiberley, S.E. 2001. The measurement of use of web-based information resources: an early look at vendor-supplied data. *College & research libraries*, 62(5): 434-453.

Caldwell, T. 2006a. ScholarlyStats won't push Sushi off the usage menu. *Information world review*, (221): 4.

Caldwell, T. 2006b. SUSHI to satisfy usage demands. *Information world review*, (220): 3.

Chandler, A. & Jewel, T. 2006a. Standards – Libraries, data providers, and SUSHI: the Standardized Usage Statistics Harvesting Initiative. *Against the grain*, 18(2): 82-83.

Chandler, A. & Jewel, T. 2006b. The Standardized Usage Statistics Harvesting Initiative (SUSHI). *Serials*, 19(1): 68-70.

Committee for Higher Education Librarians of South Africa (CHELSA), Quality Assurance Subcommittee. 2006. *Measures for quality in South African Higher Education Institution libraries. Draft 3*. [Online]. Available: http://www.ched.uct.ac.za/cil/dils/resources.html [2007, December 20]

Conyers, A. 2004. E-measures: developing statistical measures for electronic information services. *VINE: the journal of information and knowledge management systems*, 34(4): 148-153.

Coombs, K. 2005. Lessons learned from analyzing library database usage data. *Library hi tech*, 23(4): 598-609.

COSALC. 2002. Communiqué on the meeting of 31 July 2002. [Online]. Available: http://www.cosalc.ac.za/communique.html [2007, December 4]

COSALC. 2007. General communiqué, March 2007. [Online]. Available: http://www.cosalc.ac.za/communique.html [2007, December 4]

COUNTER. 2005a. The COUNTER Code of Practice for Journals and Databases. Release 2. [Online]. http://www.projectcounter.org/r2/COUNTER_COP_Release_2.pdf [2008, January 29].

COUNTER. 2005b. The COUNTER Code of Practice for Journals and Databases. Release 2. Appendix A. Glossary of terms. [Online]. http://www.projectcounter.org/r2/Appendix_A_Glossary.pdf [2008, January 17]

COUNTER. 2005c. The COUNTER Code of Practice for Journals and Databases. Release 2. Appendix H. Optional Additional Usage Reports. [Online]. http://projectcounter.org/r2/R2_Appendix_H_Jan_08.doc [2008, January 30]

Cox, J. 2003/2004. Don't confuse price with value – In academic publishing, electronic is better. *Against the grain*, 15(6): 92-93.

Davis, P. 2004. For electronic journals, total downloads can predict number of users. P*ortal: libraries and the academy*, 4(3): 379-392.

De Jager, K. 2007. Opening the library catalogue up to the web: a view from South Africa. *Information development*, 23(1): 48-53.

Emery, J. 2005. Beginning to see the light: developing a discourse for electronic resource management. *The serials librarian*, 47(4): 137-147.

Ferguson, A.W. 2003. Digital information use statistics: useful? Useless?, In *Charleston Conference Proceedings 2002*. R. Bazirjian & V., Eds. Westport, Conn: Libraries Unlimited. 27-35.

Franklin, B. 2005. Managing the electronic collection with cost per use data. *IFLA journal*, 31(3): 241-248.

Franco, T. 2005. The evaluation of electronic resources as a strategic factor in the decision making process: tools, critical points, feasible solutions. E-LIS. [Online]. Available: http://eprints.rclis.org/archives/00004163/ [2008, February 7].

Gallagher, J., Bauer, K. & Dollar, D.M. 2005. Evidence-based librarianship: utilizing data from all available sources to make judicious print cancellation decisions. *Library collections, acquisitions & technical services*, 29:169-179.

Gillingham, E. 2004. Impact of COUNTER Code of Practice on Vendor Usage Statistics. E-mail submitted to Listserv: liblicense-l@lists.yale.edu 18 June
[Accessed 28 June 2004]

Hannon, J. 2005. Monthly stats. [Personal e-mail to C. Dean, 2 September]. Buffalo, NY.

Helinsky, Z. 2007. Marketing to get better mileage from your e-resources. In *The Eresources management handbook.*, UK Serials Group. 104-111 [Online]. Available: http://uksg.metapress.com/link.asp?id=6TUU9N7WFL18 [2007, October 25].

Hiott, J. 1999. Making online use count. *Library journal*, 124(16): 44-47.

ICOLC. 2006. Revised guidelines for statistical measures of usage of web-based information resources [Online]. Available:
http://www.library.yale.edu/consortia/webstats06.htm [2008, January 28]

Kraemer, A. 2006. Ensuring consistent usage statistics, Part 2: working with use data for electronic journals. *The serials librarian*, 50(1/2): 163-172.

Lichtenberg, J. 2004. Industry investigating how to use 'usage statistics'. *Publisher's weekly*, 251(1): 12.

Lingle, V.A. 2005. Implementing EBSCO's A-to-Z and LinkSource products for improved electronic journal management, In *Electronic journal management systems: experiences from the field.* G. Ives, Ed. Haworth Press, 2005: 43-54.

McClure, C.R. 2004. Strategies for collecting networked statistics: practical suggestions. *Vine: The journal of information and knowledge management systems*, 34(4): 166-171.

Medeiros, N. 2005. Electronic resource usage statistics: the challenge and the promise. *OCLC Systems and services*, 21(3):145-147.

Moen, W.E, Oguz, F. & McClure, C.R. 2004. The challenges of nonstandardized vendor usage data in a statewide metasearch environment: the library of Texas experience. *Library quarterly*, 74(4): 403(20). [Online] Available: http://infotrac.galegroup.com [2008, February 7].

Nisonger, T.E. 2000. Usage statistics for the evaluation of electronic resources. *Library collections, acquisitions & technical services*, 24: 299-302.

Pesch, O. 2004b. Usage statistics: Taking e-metrics to the next level. *The serials librarian*, 46(1/2):143-154.

Pesch, O. 2004a. What is COUNTER and what does it mean to you? *At your service: an EBSCO publication for serials librarians & information professionals*, (68): 4-6.

Pesch, O. 2006. Ensuring consistent usage statistics, Part 1: Project COUNTER. *The serials librarian*, 50(1/2): 147-161.

Peters, T.A. 2002. What's the use? The value of e-resource usage statistics. *New library world*, 103 (1172/1173): 39-47.

ProQuest. 2007. ProQuest technical notice – usage system. [E-mail to ProQuest customer mailing list, ProQuestmail99@il.proquest.com, 18 December] [2007, December 18]

Renwick, S. 2005. Knowledge and use of electronic information resources by medical sciences faculty at the University of the West Indies. *Journal of the Medical Library Association,* 93(1): 21-31.

Sack, J. 2003/2004. The beginning of value assessment: usage information in the e-journal age. *Against the grain,* 15(6): 36-40.

Sadler, K. 2007. UKSG Seminar – Usage Statistics Training, Dublin 15 January. [E-mail to Lis-e-journals@jiscmail.ac.uk, 18 December] [2007, December 18]

SASLI. 2006. South African Site Licensing Initiative Agreements: Complete list 2003-2005. [Microsoft Excel spreadsheet e-mailed to SASLI contacts by Susan Veldsman, 18 August, with subject heading: Renewals 2007: Report back: Number One]. (Unpublished)

Schmidt, J.K. 2006. CDL-Mid-Winter discussion (#2): Use statistics for electronic resources.). E-mail submitted to Listserv: colldv-l@usc.edu 16 February [2006, February 16].

Shepherd, P.T. 2007a. Final report on the investigation into the feasibility of developing and implementing journal usage factors. [Online]. Available: http://www.uksg.org/sites/uksg.org/files/FinalReportUsageFactorProject.pdf [2008, January 30].

Shepherd, P.T. 2007b. International dateline – COUNTER-based Journal Usage Factor: a meaningful new measure? *Against the grain,* 19(4):75.

Shepherd, P.T. 2006. COUNTER: usage statistics for performance measurement. *Performance measurement and metrics*, 7(3): 142-152.

Soule, S. 2006. Cambridge University Press Journals now compliant with COUNTER 2. E-mails submitted to Listserv: liblicense-l@lists.yale.edu 5 March. [2006, March 6].

South African National Library and Information Consortium. [2007a]. About SANLiC. [Online]. Available: http://www.cosalc.ac.za/home.html. [2007, December 4].

South African National Library and Information Consortium. [2007b]. Consortium Members. [Online]. Available: http://www.cosalc.ac.za/members.html. [2007, December 4].

South African National Library and Information Consortium. [2007c]. SANLiC Manager. [Online]. Available: http://www.cosalc.ac.za/communique.html. [2007, December 4].

Stubbings, R. & Hamblin, Y. 2004. How to make your e-resources earn their keep. *Serials*, 17(1): 25-30.

Taylor-Roe, J. & Spencer, C. 2005. A librarian's view of usage metrics: through a glass darkly? *Serials*, 18(2): 124-131

Tenopir, C. 2003. Predicting the future of databases. *Library journal*, 128(10): 33.

Thomas, G. 2007. Academic libraries in the second decade of democracy: critical issues and challenges. In *Libraries for the future: progress and development of South African Libraries*, T. Bothma, P. Underwood and P. Ngulube, Eds. Pretoria: Library and Information Association of South Africa. 71-87.

White, A. & Kamal, E.D. 2006. *E-metrics for library and information professionals*. London: Facet Publishing.

Wisniewski, J. & Fichter, D. 2007. Electronic resources won't sell themselves: marketing tips. *Online*, 31(1): 54-57.

Appendix: Survey of South African academic institution libraries in the SASLI consortium

Dear Colleague

I would appreciate it greatly if you would answer the following questions with regards to electronic resources statistics in your institution. This survey forms part of my dissertation in partial completion of my MBibl (from UCT) and I am collecting responses from Higher Education libraries.

Once the results have been collected and analysed I hope to be able to use the information to establish a benchmark for electronic resources statistics in South African Academic Libraries. Please note that individual institutions will not be identified in the final document.

This questionnaire has been structured as a Word document to allow you to insert your answers and to write as much as you like in answering the various questions.

Please e-mail the completed form to me at caroline@uctlib.uct.ac.za, or fax it to me at 021 6851734, before 24 November 2006.

Thank you,
Caroline Dean
Electronic Resources Librarian
University of Cape Town Libraries

Statistics for electronic resources

Please supply the following information for administrative purposes:

YOUR NAME:

YOUR INSTITUTION:

YOUR POSITION:

QUESTION 1

What publisher/vendor produced usage statistics for electronic **databases** does your institution collect and keep? Check all that apply.

a) Number of searches

b) Number of sessions

c) Number of fulltext downloads

d) Number of turnaways

e) Please list any other database usage statistics that you keep.

QUESTION 2

What publisher/vendor produced usage statistics for electronic **journals** does your institution collect and keep? Check all that apply.

a) Number of searches

b) Number of sessions

c) Number of fulltext downloads

d) Number of turnaways

e) Please list any other electronic journal usage statistics that you keep.

QUESTION 3

Do you keep statistics for electronic books? If yes, what statistics do you keep?

QUESTION 4

What other figures do you keep? Check all that apply:

a) Number of database subscriptions

b) Number of electronic journal subscriptions

c) Number of electronic books purchased

d) Expenditure on electronic resources

e) Do you keep any other statistics for your electronic resources? Please list them.

QUESTION 5

In the process of collecting your electronic resource statistics ...

a) How many staff are involved?

b) What level of staff are they?

c) How much time do they spend collecting and managing statistics for electronic resources?

d) How are the usage statistics recorded and managed? Please explain.

QUESTION 6

Do you use statistics on electronic resources? Yes or No.

If YES, do you use statistics to:

a) demonstrate level of use?

b) demonstrate costs, eg. cost per fulltext download, cost per session, cost per search?

c) determine correct subscription level in terms of number of simultaneous users?

d) monitor number of turnaways?

e) For what other purposes do you use electronic resource statistics? Please explain.

QUESTION 7

Are you aware of the Project COUNTER Initiative? How important is it for your purposes that usage statistics are COUNTER compliant? Please explain.

QUESTION 8

What other considerations/issues would you like to raise with regards to electronic resources usage statistics?

THANK YOU VERY MUCH FOR YOUR COOPERATION.